

# An Efficient Feature Reduction Approach for Heart Disorder using Neural Network

Shaik Anisur Rahiman

Dept of Computer Science, Sri Venkateswara University, Tirupati

**Abstract**— Anticipating coronary illness is viewed as one of the most troublesome difficulties in the medical care calling. With the speedy headway of advances and the utilization of various AI techniques of late, the progression of coronary illness insightful gathering has gotten progressively perceptive and exact. Consequently, improvement of AI methodologies, which can enough separate heart contamination request, is basic. This paper deals with the turn of events and planning of a brain network association for heart contamination end taking into account patients' secondary effects and causative living creatures. This examination is directed to bunch the kind of heart ailment in two unmistakable classes like consolidate present and missing. The assessment is done on the dataset taken from the University of California at Irvine Machine Learning Data Repository. The dataset contains a huge volume of feature estimations which are diminished using SVM-RFE based component assurance strategy. The dataset contains a huge rundown of capacities which is diminished using a superior part decision procedure named as covering method. The proposed covering strategy depends on an MLP-RFE computation to pick the primary highlights from the given dataset. The picked subset of elements then goes through a preprocessing step to introduce a consistency in the allocation of data. Since MLP is seen to enjoy the benefit of giving a famous execution in portrayal stage.

## I. INTRODUCTION

Coronary illness is the main source of death on the planet, representing almost 31% of all passings. Early location and treatment of numerous cardiovascular problems are exceptionally difficult, particularly in unfortunate nations, because of an absence of analytic focuses, gifted specialists, and different assets that influence the appropriate guess of coronary illness [1]. Some pervasive gamble factors, for example, diabetes, hypertension, and exorbitant cholesterol, make it challenging to recognize coronary illness. Fundamental problems instigate sporadic heart rhythms and breathing challenges, for example, aspiratory breaks, further developed jugular vein weight, and marginal edema [2]. Since the side effects of cardiovascular sickness are so differed, they should be treated with intense watchfulness. Inability to do so may adversely affect the heart [3].

The fast headway of PC advancement in present numerous years, the usage of AI development expects a critical part in the examination of heart contaminations. This paper revolves around the purposes of brain network associations in heart contamination. The clinical informatics field created around the development, the taking care of and the treatment of clinical information for various purposes.

One of these plans is to cultivate estimations prepared to make assumptions with respect to the end, the treatment or the clinical improvement of a patient. In this paper, we have proposed a useful system, which can bunch the kind of heart ailment in two extraordinary classes.

The goals of this paper are:

- To concentrate on different elements of Statlog (Heart) dataset
- To apply the SVM Recursive Feature Elimination (SVM-RFE) for lessening the quantity of characteristics
- To characterize assaults utilizing Multilayer Perceptron (MLP)

## II. FEATURE SELECTION

Include assurance methodologies decrease the dimensionality of feature space, dispense with redundant, irrelevant or noisy data. It brings the brief effects for application: speeding up a data mining computation, further developing the data quality

and the display of data mining and growing the understandability of the mining results [4] [5][6]. highlight assurance has been a working investigation district in data mining networks since it allows basically working on the understandability of the ensuing classifier models [7]. It contains to pick a subset of information factors from a dataset with particularly immense of properties by taking out features with practically no farsighted information.

Highlight decision techniques have their most outrageous significance in data mining, AI, and model affirmation, especially for gigantic datasets. The principal point of these methods is to kill unnecessary or dull features from the dataset. Feature assurance techniques have two classes: covering and channel [8][9]. The covering surveys and picks credits subject to precision measures by the goal learning computation. Using a particular learning estimation, covering on a very basic level glance through the part space by barring a couple of elements and testing the impact of component rejection on the figure estimations [10]. The component that has enormous impact in learning measure derives it does matter and should be considered as a great component.

### 2.1 SVM Recursive Feature Elimination (SVM-RFE)

SVM-RFE feature assurance strategy was proposed in [5] to lead quality decision for sickness portrayal. Settled subsets of features are picked in a continuous in invert end way, what starts with all the part factors and disposes of every single component variable. At every movement, the coefficients of the weight vector of a straight SVM are used to deal with the part situating score. SVM-RFE procedure positions all of the features according to some score work and kills no less than one features with the most insignificant scores. This cycle is repeated until the most raised portrayal accuracy is gained. Due to its actually use in picking valuable characteristics for dangerous development request, SVM-RFE obtained an uncommon universality and is remarkable as perhaps the best component assurance technique [9]. In any case, the SVM-RFE is an unquenchable strategy that solitary longings to find the best blend for portrayal.

The SVM - RFE computation [5] can be broken into four phases:

1. Train a MLP on the arrangement set;
2. Demand features using the heaps of the resulting classifier;
3. Crash features with the smallest weight;
4. Go over the cycle with the arrangement set restricted to the abundance features.

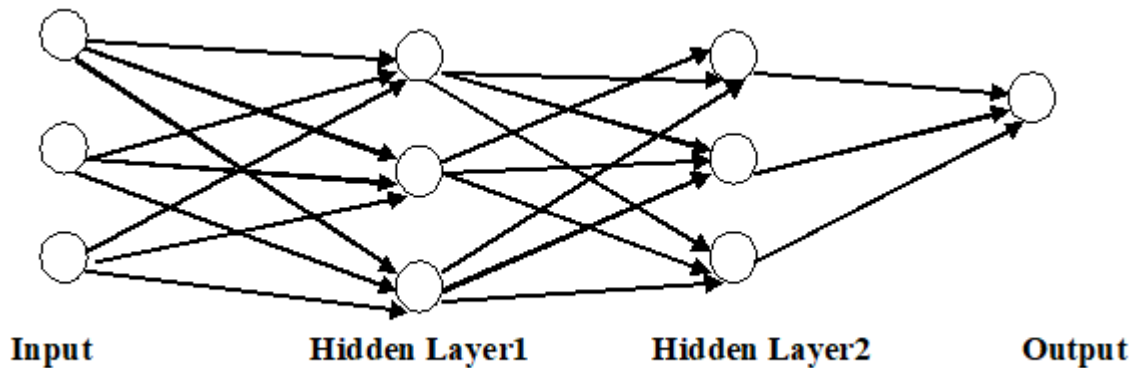
## III. METHODOLOGY

The system presented here was used a multi-layer feed-forward fake brain association was picked for this structure; it was ready in a directed way, using the backpropagation estimation.

### 3.1 Artificial Neural Network (ANN)

An ANN is a data preparing viewpoint that is moved by the way wherein a trademark material system in human mind works. ANNs are utilized completely for the strategy of various issues, including depiction, vision, talk, plan attestation, control frameworks, and so forth. A titanic number of neurons present in the human cerebrum structures the indispensable section of the brain system viewpoint and go presumably as straightforward managing portions [8][10][11]. These neurons are essentially interconnected and work in relationship to deal with complex issues. A phony neuron is a little arranging unit and plays out a reasonable calculation that is vital for the activity of a brain structure. The model of a neuron contains the pivotal parts like wellsprings of information, synaptic weights, tendency, adding crossing point, and initiation work.

According to a client perspective, ANN can be divided two immense classes dependent upon their alliance geology: feed forward and take care of in reverse brain systems. Feed forward brain systems engage the sign to stream the forward way from a certain point of view. The sign from any neuron doesn't stream to another neuron in the principal layer. In feed in reverse brain systems the sign from a neuron in a layer can stream to anything that other neuron whether it could go already or succeeding layers. Figure-1 portrays the development of this class of brain arrange.



**FIGURE 1: Structure of neural network**

### 3.2 Multilayer Perceptron (MLP)

MLP is a hero among the most extensively seen Neural Network Design that has been utilized for different applications. The MLP coordinate is ordinarily made from various focus focuses or dealing with units, and it is sorted out into a development of in any event layers [4]. The fundamental layer is named as a data layer where it gets the outer data while the last layer is a yield layer where the reaction for the issue is gotten. The covered layer is the generally engaging layer in the information layer and the yield layer, and may shape with in any occasion one layer. The goal of Multilayer Perceptron learning is to track down the best loads that limit the separation between the data and the yield. The majority of arranging calculation are utilized in Neural Network is Back Propagation and it has been utilized in managing different issues in model certification and depiction [11]. This calculation relies upon several limits like Learning Rate, Momentum Rate and Activation work, etc

In this proposed structure, incorporate abatement method using SVM-RFE is coordinated as a basic development towards lessening the amount of characteristics without losing the rule reason and target information from the primary data. The accompanying stage is encouraging a marker with a superior precision to arrange instructive record. There are different stages in the proposed plan for an efficient Arrhythmia course of action. We are proposing one more model for capable component assurance and Arrhythmia assumption. This philosophy is of the going with steps as follows:

Stage 1: Read the Statlog (Heart) dataset.

Stage 2: Preprocess the dataset.

Step 2.1: Imputer the missing characteristics using mean

Stage 3: Select the tremendous features using SVM-RFE computation.

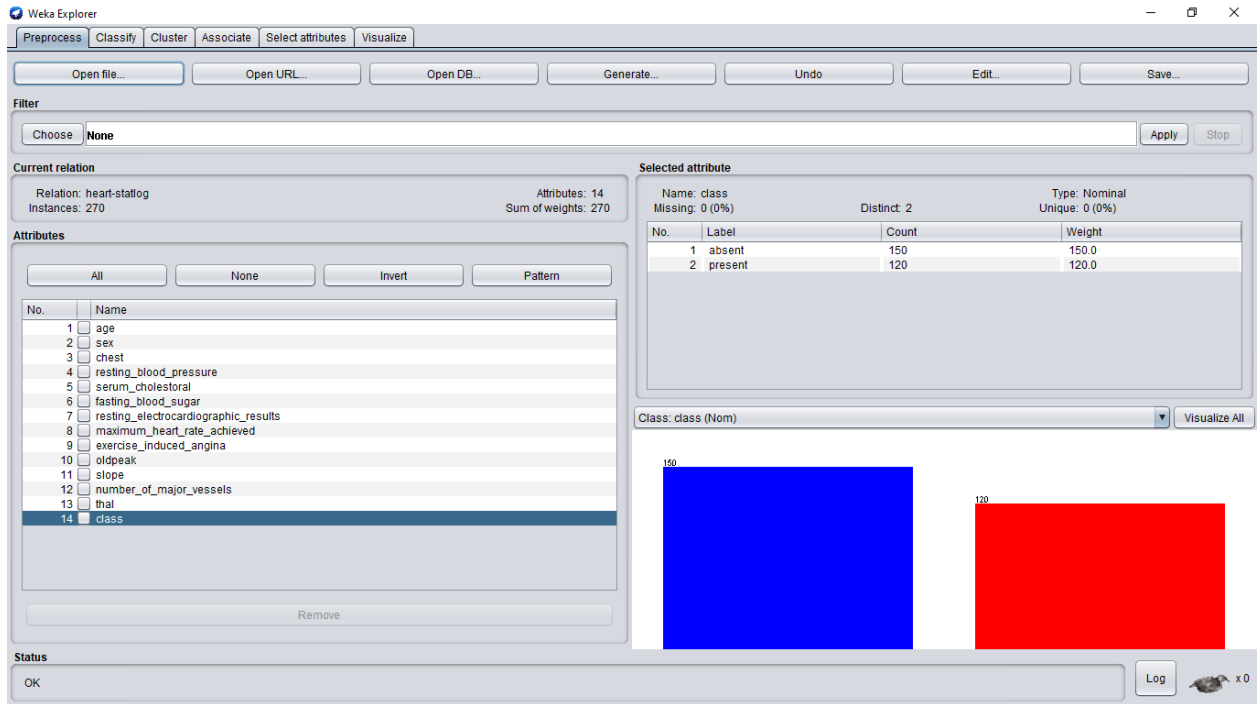
Stage 4: perform Classification using MLP estimation on the dataset to pick the best features.

Stage 5: Evaluate execution of the classifier.

## IV. EXPLORATORY DETAILS AND RESULTS

We have thought about the Statlog (Heart) dataset from the UCI Machine Learning Repository data [12] to survey execution of MLP course of action. The Statlog (Heart) dataset contains 270 cases and 13 credits, having 2 class names (150 missing and 120 present). The evaluations have been driven by using WEKA. WEKA is a best in class office for making AI (ML)

systems and their application to real information mining issues. The detailed statistical summary of the heart dataset as shown in the figure-2.

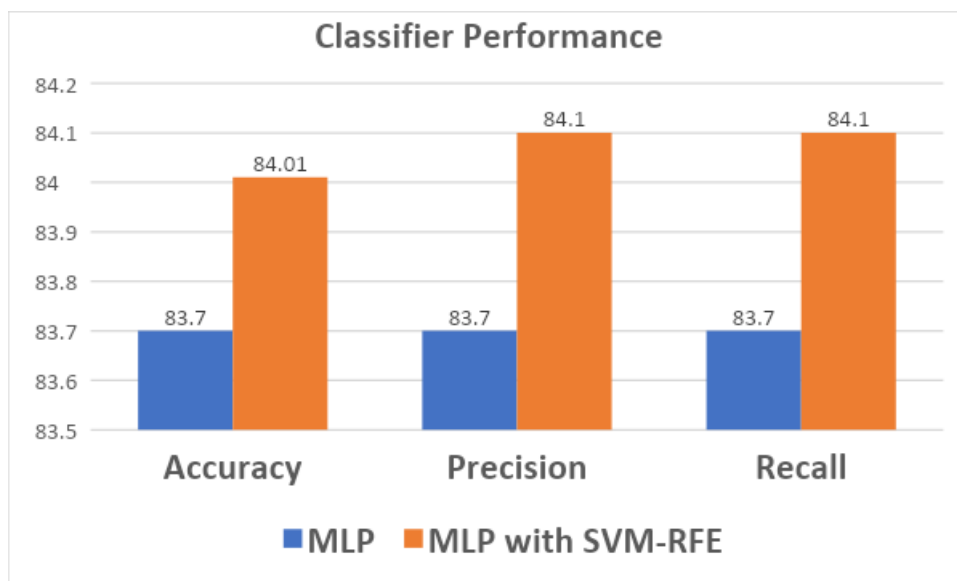


**FIGURE 2: Summary of the dataset**

As the dataset has a large set of features, feature selection is applied to select the most relevant and significant features containing useful information required for data classification. We use 70% of records as the training data and the other 30% as the testing data.

### V. RESULTS

In the first stage MLP algorithm is trained on the original set of features was used in the experiment. In the second stage we implement a SVM-RFE algorithm for obtaining the adequate number of features to identify the features selected. The results that we got for MLP algorithm for the full dimension data and also after the feature reduction with SVM-RFE technique are shown in the figure-3. The Statlog (Heart) dataset is labeled as one of 2 different heart disease types.



**FIGURE 3: Performance of MLP and MLP with SVM-RFE**

From the figure-3, we observe the performance of MLP without SVM-RFE based on accuracy has got 83.7%, whereas the performance of MLP with SVM-RFE feature selection based on accuracy has achieved 84.01%. However, there is an improvement in the accuracy with feature selection. The accuracy rate is increased 0.31% with feature selection. The experimental screen shots are shown in the figure-4 and figure-5.

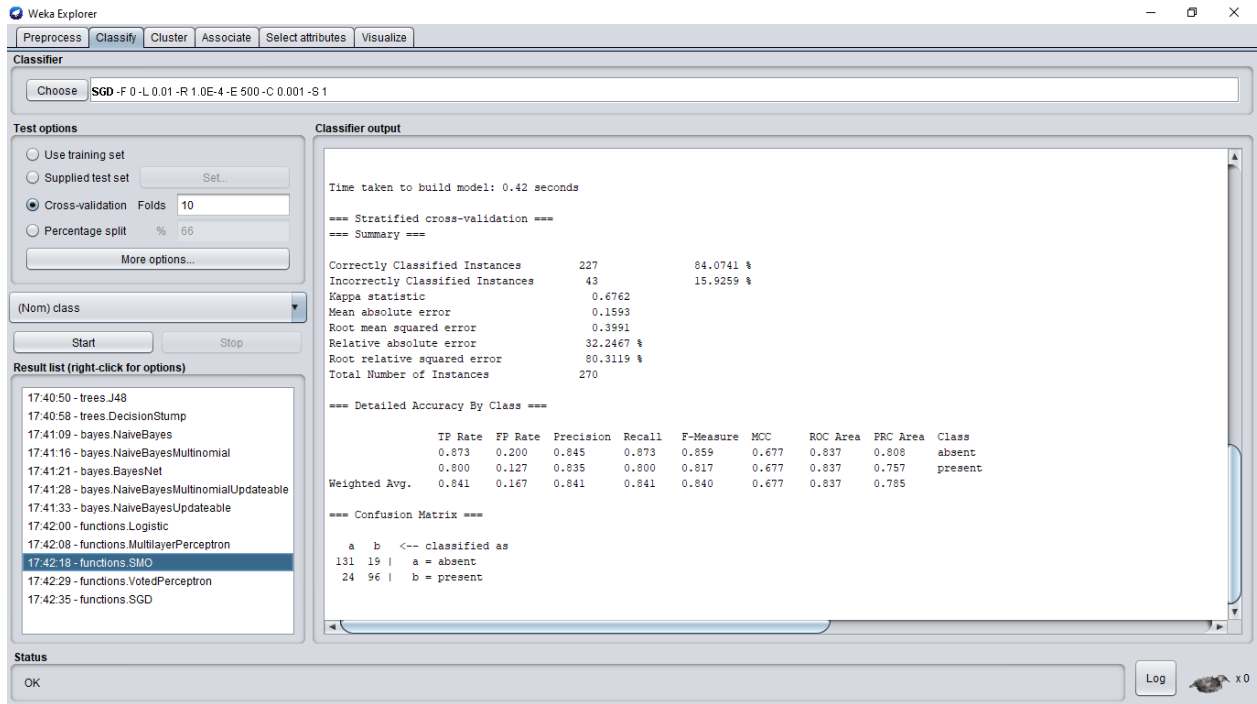


FIGURE 4: Results Screen shots

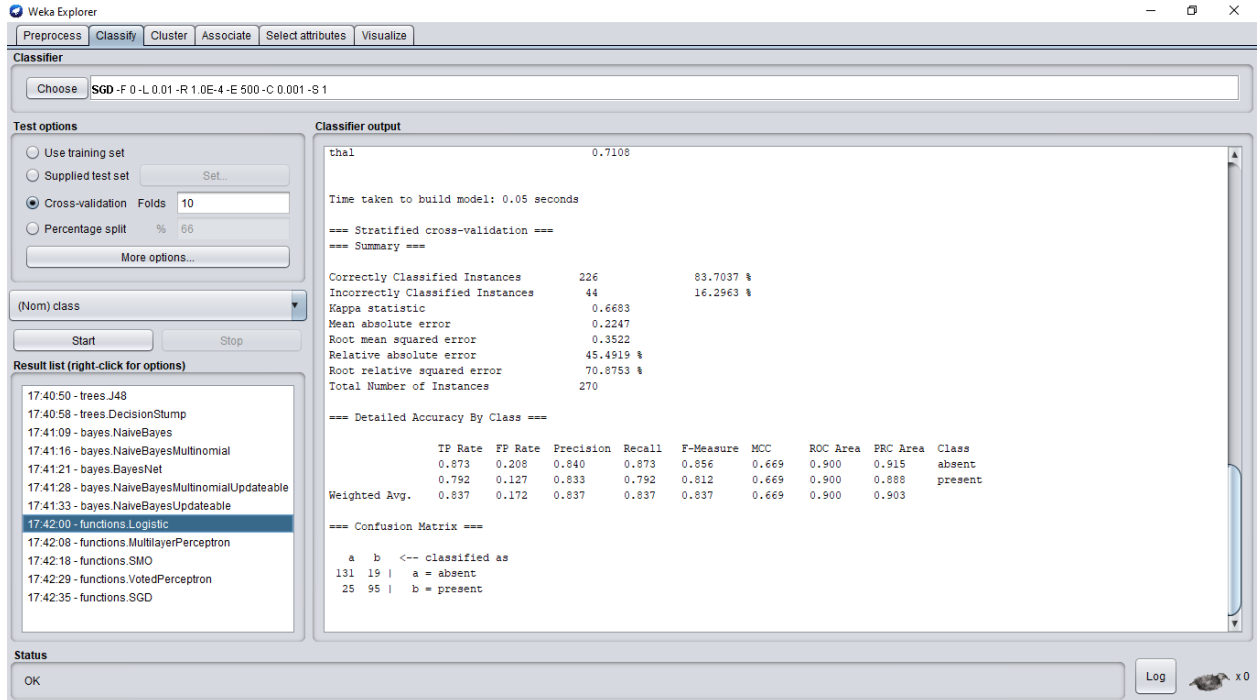


FIGURE 5: Results Screen shots

In our exploratory outcome the MLP with SVM-RFE calculation shows the most noteworthy exactness contrasted and MLP without SVM-RFE. With the improvement the precision, the proposed model exhibited that it performs well subsequent to choosing pertinent highlights. This outcome gave new understanding utilizing a characterization learning calculation and decrease method to choose applicable and significant component to work on the exactness of the framework and to recognize

potential highlights which might add to this improvement. The vast majority of the proposed research framework could really use highlight choice cycle to further develop recognition pace of their framework and limit significantly the misleading problem rate.

## VI. CONCLUSION

In this paper, the significance of using a lot of relevant features with the MLP game plan learning estimation for Statlog (Heart) disorder figure has been represented. A show and idea of a component assurance methodology which contain a recursive part removal using a MLP classifier to recognize huge features have been done. The part decision, preprocessing, and course of action procedures have conveyed a mix which gives promising results to disorder gathering. The appraisal the feasibility of the procedure using particular plan metric assessment has been made and it has been shown that by diminishing the number of features, the accuracy of the model was gotten to the next level. To perceive Statlog (Heart) disorder from tremendous dataset, acknowledgment computation, and feature assurance procedure have unreasonably more capable.

## REFERENCES

- [1] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019.
- [2] B. Brahmī and M. H. Shirvani, "Prediction and diagnosis of heart disease by data mining techniques," *J. Multidisciplinary Eng. Sci. Technol.*, vol. 2, no. 2, pp. 164–168, Feb. 2015.
- [3] C. Yang, B. An, and S. Yin, "Heart-disease diagnosis via support vector machine-based approaches," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 3153–3158.
- [4] Dr. G. Ravi Kumar and Dr. K. Nagamani, "Banknote Authentication System utilizing Deep Neural Network with PCA and LDA Machine Learning Techniques", *International Journal of Recent Scientific Research*, Volume-9, Issue:12(D), PP:30036-30038, 2018
- [5] Guyon, Weston, Barnhill, and Vapnik, "Gene selection for cancer classification using support vector machines," *MACHLEARN: Machine Learning*, vol. 46, (2002).
- [6] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering", *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, (2005), pp. 491–502.
- [7] H. Liu, J. Sun, L. Liu, and H. Zhang, "Feature selection with dynamic mutual information," *Pattern Recognition*, vol. 42, no. 7, pp. 1330–1339, 2009.
- [8] H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with Java implementations", San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., (2000)
- [9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, (2003) March, pp. 1157–1182
- [10] J. Han and M. Kamber, "Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.
- [11] S.Rahamat Basha and G.Ravi Kumar Surya Bhupal RaoG, "A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques", *Journal of Mechanics of Continua and Mathematical Sciences*, Special Issue, No.-5,PP:120-131, 2020
- [12] UCI machine learning repository. <http://archive.ics.uci.edu/ml/>